ABSTRACT

An author-oriented document summarizer for a word processor is described. The document summarizer performs a statistical analysis to generate a list of ranked sentences for consideration in the summary. The summarizer counts how frequently content words appear in a document and produces a table correlating the content words with their corresponding frequency counts. Phrase compression techniques are used to produce more accurate counts of repeatedly used phrases. A sentence score for each sentence is derived by summing the frequency counts of the content words in a sentence and dividing that tally by the number of the content words in the sentence. The sentences are then ranked in order of their sentence scores. Concurrent with the statistical analysis, during the same pass through the document the summarizer performs a cue-phrase analysis to weed out sentences with words or phrases that have been pre-identified as potential problem phrases. The cue-phrase analysis compares sentence phrases with a pre-compiled list of words and phrases and sets conditions on whether the sentences containing them can be used in the summary. Following the cue-phrase analysis, the summarizer creates a summary containing the higher ranked sentences. The summary may also include a conditioned sentence if the conditions established for inclusion of the sentence have been satisfied. The summarizer then inserts the sentence at the beginning of the document before the start of the text.